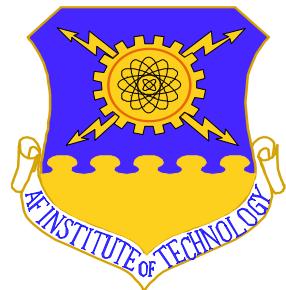


AFIT/EN-TR-02-06
TECHNICAL REPORT
April 2002



Application of Decision Analysis to Automatic Target Recognition Programmatic Decisions

William K. Klimack, Col, USA
Christopher B. Bassham, Capt, USAF
Dr. Kenneth W. Bauer, Jr.

**GRADUATE SCHOOL OF ENGINEERING AND MANAGEMENT
AIR FORCE INSTITUTE OF TECHNOLOGY
WRIGHT-PATTERSON AIR FORCE BASE, OHIO**

Approved for Public Release; Distribution Unlimited

Report Documentation Page

Report Date 29 Apr 2002	Report Type Final	Dates Covered (from... to) 01 Jul 2001 - 01 Apr 2002
Title and Subtitle Application of Decision Analysis to Automatic Target Recognition Programmatic Decisions		Contract Number
		Grant Number
		Program Element Number
Author(s) William K. Klimack, Col, USA Christopher B. Bassham, Capt, USAF Dr. Kenneth W. Bauer, Jr., Professor		Project Number ENR# 2001-018
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) Air Force Institute of Technology AFIT/ENS 2950 P Street, Building 640 Department of Operational Sciences WPAFB OH 45433		Performing Organization Report Number AFIT/EN-TR-02-06
		Sponsor/Monitor's Acronym(s)
Sponsoring/Monitoring Agency Name(s) and Address(es) Air Force Research Laboratory Sensors Directorate 2241 Avionics Circle Building 620, Rm N3-X19 WPAFB OH 45433		Sponsor/Monitor's Report Number(s)
		Distribution/Availability Statement Approved for public release, distribution unlimited
Supplementary Notes The original document contains color images.		

Abstract

The purpose of this research is to demonstrate the application of decision analysis (DA) techniques to the decisions made throughout the lifecycle of Automatic Target Recognition (ATR) technology development. This work is accomplished in the hopes of improving the means by which ATR technologies are evaluated. The first step in this research was to create a flexible decision analysis framework that could be applied to a variety of decisions across several different ATR programs evaluated by the Comprehensive ATR Scientific Evaluation (COMPASE) Center. For the purpose of this research, a single COMPASE Center representative provided the value, utility, and preference functions for the DA framework. The DA framework employs performance measures collected during ATR classification system (CS) testing to calculate value and utility scores. The authors gathered data from the Moving and Stationary Target Acquisition and Recognition (MSTAR) program to demonstrate how the decision framework could be used to evaluate three different ATR CSs. A decision-maker may use the resultant scores to gain insight into any of the decisions that occur throughout the lifecycle of ATR technologies. Additionally, a means of evaluating ATR CS self-assessment ability is presented.

Subject Terms

Automatic Target Recognition, Decision Analysis, Competing Classifiers, Measures of Performance, Utility, Value, Comparison, Value Hierarchy, Self-Assessment, Evaluation, ROC Curve, MSTAR

Report Classification unclassified	Classification of this page unclassified
Classification of Abstract unclassified	Limitation of Abstract UU
Number of Pages 38	

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

Abstract

The purpose of this research is to demonstrate the application of decision analysis (DA) techniques to the decisions made throughout the lifecycle of Automatic Target Recognition (ATR) technology development. This work is accomplished in the hopes of improving the means by which ATR technologies are evaluated. The first step in this research was to create a flexible decision analysis framework that could be applied to a variety of decisions across several different ATR programs evaluated by the Comprehensive ATR Scientific Evaluation (COMPASE) Center of the Air Force Research Laboratory (AFRL). For the purposes of this research, a single COMPASE Center representative provided the value, utility, and preference functions for the DA framework through elicitation meetings with the authors. The DA framework employs performance measures collected during ATR classification system (CS) testing to calculate value and utility scores. The authors gathered data from the Moving and Stationary Target Acquisition and Recognition (MSTAR) program to demonstrate how the decision framework could be used to evaluate three different ATR CSs. A decision-maker may use the resultant scores to gain insight into any of the decisions that occur throughout the lifecycle of ATR technologies. Additionally, a means of evaluating ATR CS self-assessment ability is presented. This represents a new criterion that emerged from this study, and no present evaluation metric is known.

Table of Contents

Title Page.....	i
Disclaimer	ii
Abstract	iii
Table of Contents.....	iv
List of Figures.....	v
List of Tables.....	vi
1.0 Summary.....	1
1.1 Introduction.....	2
2.0 Methods, Assumptions, and Procedures.....	3
2.1 The Decision Situation.....	3
2.2 Encoding the Value Hierarchy	4
2.3 Single Dimensional Value and Utility Function Elicitation	13
2.4 ATR CS Alternatives.....	15
3.0 Results and Discussion.....	16
4.0 Conclusions	22
5.0 Recommendations	23
References	24
Appendix A – Elicitation Results	25
List of Acronyms	31

List of Figures

Figure 2.1 Abstract Depiction of Target Truth and Clutter Areas.....	5
Figure 2.2 Sample Receiver Operating Characteristic (ROC) Curve	8
Figure 2.3 The Value Hierarchy for ATR Decisions.....	13
Figure 3.1 Radar Plots of ATR Value and Utility	18
Figure 3.2 Risk Profile for CID.....	19
Figure 3.3 Risk Profile for ISR.....	19
Figure 3.4 Sensitivity Analysis for CID Scenarios.....	20
Figure 3.5 Sensitivity Analysis for ISR Scenarios	20
Figure 3.6 ATR Value (With Certainty) Versus Redeployment Cost	21
Figure 3.7 ATR Value (With Uncertainty) Versus Redeployment Cost	22
Figure 3.8 ATR Utility (With Uncertainty) Versus Redeployment Cost	22

List of Tables

Table 3.1 ATR CS Expected Value and Expected Utility Results 16

Table 3.2 Recommended ATR CS Alternative by Expected Value and Utility .. 17

Application of Decision Analysis to Automatic Target Recognition Programmatic Decisions

1.0 Summary. The purpose of this research is to demonstrate the implementation of a decision analysis (DA) approach towards programmatic decision-making within the Automatic Target Recognition (ATR) field of research. Several decisions within the lifecycle of ATR research are based upon evaluations using many performance measures and program characteristics. Often, the evaluation of these raw performance measures and characteristics leads to non-dominating solutions among the various ATR programs and technologies competing for further development. Therefore, the decisions are finalized using the preferences and values of the decision-maker. Thus, it is clear that ATR research decision-makers should employ a formalized decision analysis framework to aid in selecting the best ATR research direction or ATR product available based upon the given performance characteristics of ATR classification systems (CSs) and to ensure evaluation fairness among competing technologies. The DA framework presented quantifies and rank-orders the preferences of the decision-maker, i.e., the desirable traits of a good ATR CS are weighted more heavily than less desirable traits. The numerous performance measures and characteristics of a given set of ATR CSs may then be synthesized through the DA model. The result is a single utility, or value, measure associated with each ATR CS under consideration. This methodology provides the decision-maker with a defensible approach that includes his own value structure of the problem for use in making programmatic decisions. This research uses past ATR data to illustrate that this technique is feasible and potentially very useful to the decision-makers involved in ATR research.

1.1 Introduction. Automatic Target Recognition is a processing problem where an image is examined in order to detect and classify objects of interest, or *targets*. The image is provided by one or more sensors, which typically are forward-looking infrared, millimeter wave, synthetic aperture radar, or laser radar systems. An ATR CS, in the form of a pattern recognition and classification algorithm, is then applied to the image in order to identify particular regions of interest (ROIs) and to then classify whether the ROI is a target or not. It is necessary to point out that both enemy and friendly objects of interest are referred to as targets. Typically, the targets are difficult to separate both from normal environmental objects, generally referred to as *clutter*, and objects with target-like signatures found within the image, often referred to in ATR research as *confusers*. Currently, human analysis far exceeds the capabilities of automated ATR systems. It is highly desirable to improve automated ATR capabilities, which would increase analytic capacity in military intelligence systems as well as permit ATR systems to be employed on unmanned platforms. Automatic Target Recognition is widely acknowledged as a critical military capability [4].

ATR CSs generally fall into three classes: statistical pattern recognition, neural networks, or model-based recognition. All ATRs have a number of quantifiable evaluation measures, such as ATR CS performance, robustness, estimate accuracy, employment doctrine, and cost, which may be used to compare multiple ATR CSs. In general, these measures are not assessed in total, but specific measures are selected when considering decisions for a specific program. No examples of employment of DA techniques with respect to ATR selection have been found in the literature.

2.0 Methods, Assumptions, and Procedures. The following section details the steps taken to construct a DA framework. The first step is defining the decision situation with a decision-maker. Next, the DA model is generated using quantitative representations of the preferences held by the decision-maker. Finally, the performance measures of multiple ATR CSs may then be introduced to the model for analysis.

2.1 The Decision Situation. The objective of ATR evaluation is to determine which system performs best via performance measure assessment and comparison. During development of past ATR CSs, decisions that compared competing ATR CSs were not performed in a manner that collectively considered all pertinent aspects of the decision. Typically a subset of criteria was examined. Employment of decision analytic techniques permits all criteria to be considered and trade-offs performed when comparing systems. As an example of how these techniques can be applied, the Moving and Stationary Target Acquisition and Recognition (MSTAR) program, which was an investigative effort into synthetic aperture radar (SAR), model-based ATR development, provides an ideal scenario [5]. At the time, three ATR systems were tested as candidates for furthering the development of an advanced ATR system [5]. These ATR CSs are referred to as ATR 33, ATR 55, and ATR 89.

There are two basic ATR employment profiles under consideration. The Combat Identification (CID) employment profile is implemented when the primary objective of the ATR CS is to select targets for weapon systems. In this scenario, the system is allowed to sacrifice detection performance in order to gain classification accuracy. Thus, the selection of a target must be associated with a high degree of confidence as to

minimize the number of false alarms. The Intelligence/Surveillance/Reconnaissance (ISR) employment profile is used when the primary objective of the ATR CS is to collect information for many potential targets. In this scenario, the CS is allowed to sacrifice classification accuracy for improved detection performance. Thus, the goal of the ATR CS is to detect as many potential targets as possible as to minimize the number of targets that evade detection.

In order to assist the Comprehensive ATR Scientific Evaluation (COMPASE) Center, an Air Force Research Laboratory (AFRL) organization within the Sensors Directorate, with its assessment of which future ATR systems should be retained for further development, a feasibility study of value-focused decision analysis is performed upon the MSTAR data. Dr. Timothy Ross, COMPASE Center Director, served as the decision-maker. Dr. Ross participated in two elicitation sessions. The initial session developed the subject's value hierarchy and elicited single dimensional value functions for the evaluation measures. Single dimensional utility functions were elicited during the second session as well as follow up clarification on several minor issues. Additionally, the authors discussed minor points with Dr. Ross via e-mail. Throughout the elicitation process, the subject was comfortable expressing his preferences in a quantified manner. Discussions with Dr. Ross confirmed and refined the evaluation measures for the decision situation.

2.2 Encoding the Value Hierarchy. In a test environment, ATR CS evaluation is accomplished using performance measure assessment and comparison. Assessment refers to the collection of several quantitative performance measures. Understanding the design of the testing environment is key to the understanding of performance measure

assessment. To begin, the test image is broken into two separate areas: the *target truth area* and the *clutter scene area*. Testing personnel place targets and confusers within the target truth area. The ATR CS examines this section of the scene for potential targets. Any target declaration made against a confuser, or *non-targets*, is considered a *false alarm*. Next, the ATR CS scans the clutter scene area, which is known to be devoid of targets. Any target declarations made by the ATR CS in this section are considered to be false alarms *in clutter*, since no confusers exist in this area. Figure 2.1 illustrates the manner in which an ATR test area is divided. When an ATR CS has scanned both areas, ATR performance may then be assessed.

Target Truth Area	Clutter Area
<p>Key: T → Target N → Confuser C → Clutter □ → Declared as Target</p>	<p>Key: T → Target N → Confuser C → Clutter □ → Declared as Target</p>
<p>Figure 2.1 Abstract Depiction of Target Truth and Clutter Areas.</p>	

The primary performance measure is the probability of detection, P_D . This measure is defined as:

$$P_D = N_{DET} / N_{TGT} \quad (1)$$

where N_{DET} is the number of targets detected by the ATR CS and N_{TGT} is the total number of targets to be detected within the image. The probability of detection provides insight into how well the ATR CS is detecting the target it is designed to find. Beyond detecting targets, it is desirable that the ATR CS provides a conjecture of the taxon, or label, of the object. The ATR may or may not further refine the description of the target. Typically, a refined identification is bifurcated into class and type, but a target may be labeled at any level desirable to the decision-maker. *Class* describes a broad category of materiel. Example class taxa are main battle tanks (MBTs) and surface-to-air systems. The *type* of the object is the specific nomenclature. The M-1A1 and T-72 main battle tanks are ‘type’ examples of the MBT class. Given the targets detected by the ATR CS, additional performance measures conditional to P_D may then be calculated concerning how well the ATR algorithm classifies the ROIs that it considers targets. The first performance measure, the probability of correct classification (P_{CC}), is the ratio of the number of targets correctly classified by class (N_{CC}) to the total number of detected targets (N_{DET}).

Thus, P_{CC} is defined as:

$$P_{CC} = N_{CC}/N_{DET}. \quad (2)$$

In other words, P_{CC} measures the proportion of correctly labeled targets to the number of correctly detected targets, e.g., classifying a detected target as a MBT when it indeed is a MBT, and serves as an estimate of the probability of correctly classifying a target. A similar, but slightly more specific, performance measure is the probability of correct identification, or P_{ID} . This measure is the ratio of the number of targets correctly classified by type (N_{ID}) to the total number of detected targets (N_{DET}). Thus, P_{ID} is defined as:

$$P_{ID} = N_{ID}/N_{DET}. \quad (3)$$

In other words, P_{ID} measures the proportion of correctly labeling, for example, a ROI as a T-72 MBT when it indeed is a T-72 MBT.

While the previous performance measure indicate how well an ATR CS performs in detecting targets, the next set of measures provides insight into how often an ATR CS mistakes non-targets as targets. The probability of false alarm measure, denoted P_{FA} , is the ratio of the number of detected confusers, or non-targets, (N_C) to the total number of known confusers in the image (N_{FA}), and is defined as:

$$P_{FA} = N_C / N_{FA}. \quad (4)$$

A similar performance measure is the false alarm rate (FAR), which is the ratio of the number of false alarms in clutter (N_{CL}) to the clutter scene area (A), defined as:

$$FAR = N_{CL} / A. \quad (5)$$

This performance measure indicates how likely an ATR is to mistake terrain or natural objects as a potential target.

Typically, P_D is changed for an ATR algorithm by adjusting a detection threshold internal to the CS. P_D may be increased to any desired level as the detection threshold is adjusted, but there is a corresponding degradation in the ATR performance as more clutter and confusers are incorrectly declared as targets. That is, the false alarm rate (FAR) and probability of false alarm, P_{FA} , never decrease, and should increase, as the probability of detection increases. For a given ATR CS, a receiver operating characteristic (ROC) curve illustrates the trade-off between the detection performance and the false alarm rate. Figure 2.2 depicts sample ROC curves along with an area under the curve measure, which is typically used in comparing multiple ROC curves. When

evaluating ATRs, either the P_D is fixed for a particular mission profile, or the ROC curve is employed. In both cases, P_D is not considered an evaluation measure as the ATR performs at that level by definition. The false alarm performance is then the concern.

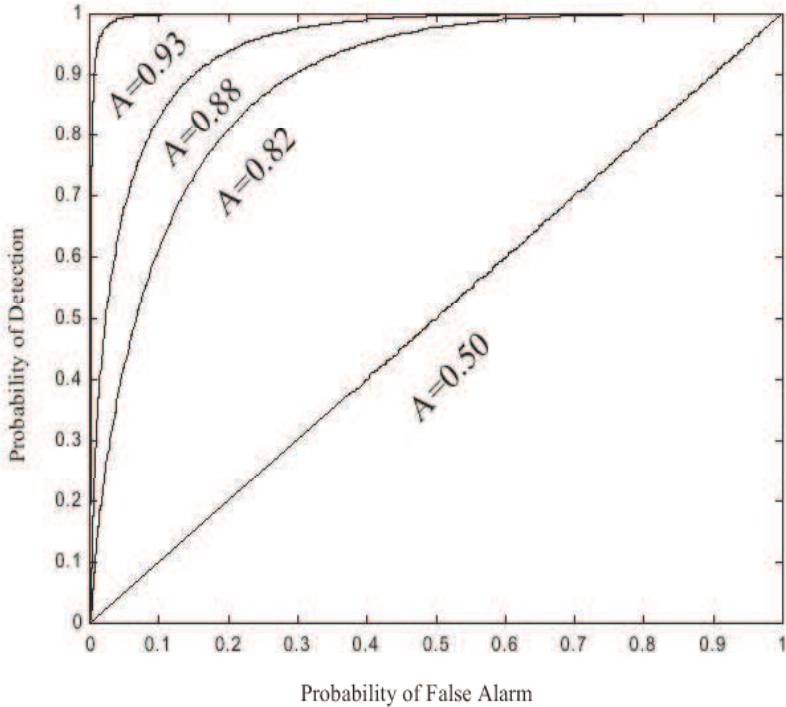


Figure 2.2 Sample Receiver Operating Characteristic (ROC) Curve [2]. The ROC curve represents the performance of a given ATR CS as an internal detection threshold is varied. The plot above illustrates the ROC curve for four different ATR CSs and provides a sample area under the curve (A) performance measure for each (higher is better).

Another desirable characteristic of an ideal ATR CS is its frequency of making declarations of targets and non-targets to be detected. Thus, a superior ATR system will declare a larger set of the known target population in an image than an inferior system. This measure hints at the confidence an ATR has in its detection ability. The probability of declaration, or P_{DEC} , is defined as:

$$P_{DEC} = N_{DEC} / N_{TOTAL} \quad (6)$$

where N_{DEC} represents the number of correct declarations (declaring a ROI to either a target or a non-target) made by the ATR CS and N_{TOTAL} is the total number of test objects, both targets and confusers. Thus, N_{DEC} consists of all ROIs declared as targets that are targets and all ROIs declared to be non-targets by the ATR CS that are confusers.

It is also desirable that ATR performance is robust to target, environmental, and sensor differences; provide an assessment of its confidence in its target estimates; and have a well-developed employment concept. ATR CSs are typically trained against a given set of baseline, or *nominal*, target images. An example of ATR performance robustness occurs when an ATR CS is able to detect and possibly correctly classify an MBT even though it has not been presented with the particular configuration that the vehicle exhibits. Thus, the ATR CS detects and/or correctly classifies the MBT when a variety of external differences from the nominal image training set, such as open hatches, turret articulation, or external stores, appear. Performance measures based upon the nominal training set, which are generally near perfect, are then compared to the performance measures where *at least one change* is made in the target configuration. It is desirable that the probabilities of detection, typification, and classification (P_D , P_{ID} , P_{CC}) remain as close to their nominal values as possible. Degradation is measured in percent change from the nominal values for some specific target set where the targets are perturbed in some fashion. The percent change values for probability of detection are calculated in the following manner, and P_{CC} and P_{ID} are assessed similarly. A nominal probability of detection measure (P_{D-NOM}) is assessed against targets at the baseline configuration. Next, a probability of detection measure is assessed against all other target

configuration deviations of interest (P_{D-DEV}), which may include sensor, target, environmental, or some combination of changes to the configuration. The difference between the two measures is assessed:

$$\Delta P_D = P_{D-NOM} - P_{D-DEV} \quad (7)$$

and the resultant difference estimate is between 0 and 1. For a large number of observations on a target array, the Central Limit Theorem permits the assumption of normality, and the formula:

$$\Phi(x) = \Delta P_D \pm Z_\alpha \sqrt{\frac{\Delta P_D(1 - \Delta P_D)}{n}} \quad (8)$$

can be used to create a probability density function, $\Phi(x)$, around the estimate for use in the decision analysis model [3]. To complete the procedure and create a percent change in the detection difference measure ($\% \Delta P_D$), the upper, lower, and mean values of the probability of detection difference estimate must be multiplied by 100. Though these estimates could be expressed as a normal distribution within the decision analysis model, they may be approximated by a triangular distribution. The triangular distribution forces the realizations to be bounded within the domain of elicitation of the DA information.

The ATR CS self-assessment is a confidence, expressed as a probability of accuracy, for the detection, typification, and classification estimates, C^D, C^{ID}, C^{CC} , respectively, as determined by the CS.. For example, an ATR CS may have difficulty with the correct identification of a target. Perhaps the target exhibits characteristics that indicate to some degree that it is a T-72 MBT, but other characteristics indicate that it is a T-80 MBT. The ATR CS provides an identification, declaring that it is either a T-72 or a T-80, and an associated confidence, C_{ID} , for its declaration. A CS may be designed to

return a single target type with an associated confidence (e.g., 90 percent confidence that the target is a T-80), or a list of several possible target types each with an associated confidence (e.g., 60 percent confidence that the target is a T-80 and 40 percent confidence that the target is a T-72). Detection and classification confidences perform similarly.

This self-confidence score is compared to the true target identity to assess the accuracy of the confidence estimate. The accuracy of these confidence estimates for detection, typification, and classification estimates (E_{S-PD} , E_{S-PCC} , E_{S-PID}) are scored on the unit interval. The inclusion of the self-assessment measure accuracy in the DA model is for use on future ATR systems with this capability. The ATR CSs of the MSTAR program did not have the capability to provide or use these measures.

One possible method of assessing E_{S-PD} , E_{S-PCC} , E_{S-PID} would be to compare C^D, C^{ID}, C^{CC} to the true nature of the target. Continuing with the identification case, each target is assigned C_{ij}^{ID} by the ATR CS, where the variable i indicates the target and the variable j indicates various identifications of the same target i by the CS. The self-assessment accuracy, E_{S-PID} , may be determined from

$$E_{S-PID} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J \left[\left(1 - \frac{\sum_{i=1}^J (1 - T_{ij})(C_{ij}^{ID})}{\sum_{i=1}^J C_{ij}^{ID}} \right) T_{ij}(C_{ij}^{ID}) \right] \quad (9)$$

where J is the total number of possible target identifications for the i th target, N is the total number of targets, M is the total number of target identifications made by the CS, and T_{ij} is an indicator variable defined as:

$$T_{ij} = \begin{cases} 1, & \text{ID correct} \\ 0, & \text{ID incorrect} \end{cases} \quad (10)$$

Similarly, E_{S-PD} , and E_{S-PCC} are defined by

$$E_{S-PD} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J \left[\left(1 - \frac{\sum_{i=1}^J (1-T_{ij})(C_{ij}^D)}{\sum_{i=1}^J C_{ij}^D} \right) T_{ij}(C_{ij}^D) \right] \quad (11)$$

and

$$E_{S-PCC} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J \left[\left(1 - \frac{\sum_{i=1}^J (1-T_{ij})(C_{ij}^{CC})}{\sum_{i=1}^J C_{ij}^{CC}} \right) T_{ij}(C_{ij}^{CC}) \right] \quad (12)$$

respectively.

Finally, cost is a consideration. Costs may be grouped by developmental costs, redeployment costs, and operational, or use, costs. Developmental costs are those incurred when bringing the ATR to an operational status. Redeployment costs are those incurred when moving the ATR from its base to an operational area. Operational costs are those incurred when employing the system. Each of these categorical costs may be further decomposed into the sub-costs of funding, time, expertise, and risk, except that there are no operational risks (with respect to the system performance).

The value hierarchy that emerged from the elicitation is depicted in Figure 2.3. The parenthetical numbers are the weights representing the relative importance, within the parent value, of the values. The value hierarchy was decomposed until the bottom tier members were measurable.

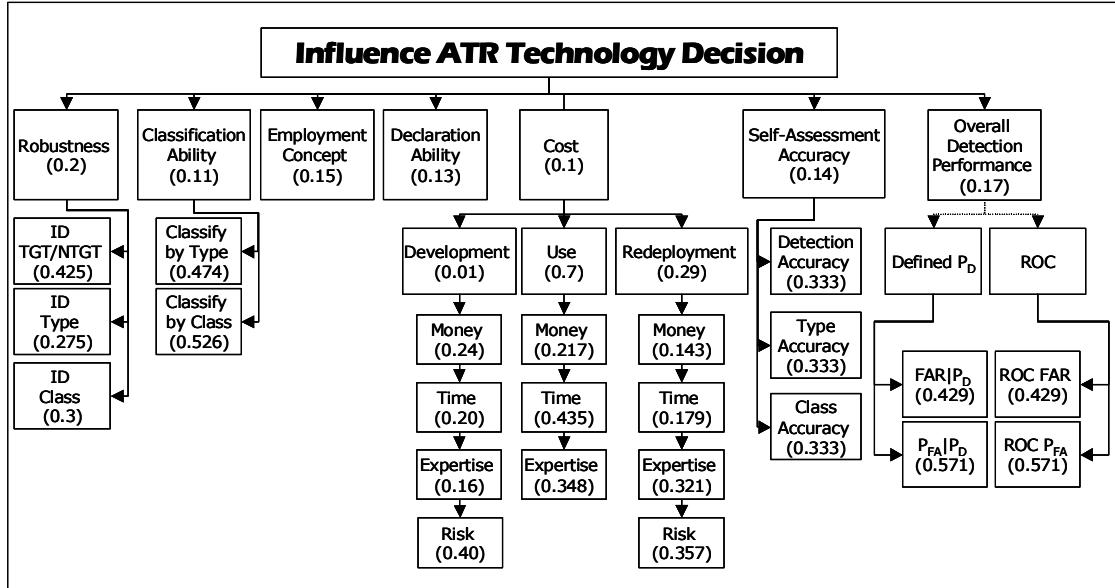


Figure 2.3 The Value Hierarchy for ATR Decisions. The dashed line under Overall Detection Performance indicates that either the Defined P_D or the ROC measures are employed. The parenthetical numbers indicate the relative weights for a value, within the parent value.

In order to assess the categorical data, scales were constructed. For the adequacy of employment concept measure, the created scale categorized the employment concept as well defined, strongly defined, moderately defined, poorly defined, and no definition. Cost risks were assessed on a scale with low, medium, and high risk. Because these scales were developed in cooperation with the subject, no precise definitions were specified for the terms.

2.3 Single Dimensional Value and Utility Function Elicitation. The single dimensional value functions were elicited for each of the evaluation measures (the bottom tier of the value hierarchy). It was explained that most and least preferred levels of an evaluation measure mapped to one and zero respectively, and the function captures intervening preference levels. Initial discussion introduced the idea of comparing relative

differences in preference between pairs of evaluation measure observations. However, the subject was very comfortable with expressing preferences in a quantitative manner, usually commenting on a functional form and providing a value to fix the curve. For example, he often observed that his preference would decrease exponentially on a measure and pass through a point he specified. Direct assessment of the value functions was employed.

Single dimensional utility functions were elicited for continuous evaluation measures employing the certainty equivalent lottery technique. In this process the subject chooses between uncertain alternatives with two equally likely outcomes, each at the extremes of the evaluation measure domain, or a certain alternative with some specified level of the evaluation measure. The certain alternative is varied until the subject is indifferent between the alternatives. As is considered standard practice, the certainty equivalent was approached by alternately providing values from the margins of the evaluation measure domain to avoid anchoring. The initial lottery considered the entire domain for the evaluation measure under consideration at that point. This provides the datum point (known as a mid-value point) for which the subject assigns a utility of 0.5. This process was repeated for the sub-domains created about the mid-value point (the fractile approach). Thus each utility function is established with five data points, three of which are elicited and two defined. The subject again was comfortable with the elicitation methodology and often quickly determined his indifference points. As no functional form was evident from these elicitations, linear interpolation was employed for utilities between elicited points.

For the evaluation measures with categorical scales, the probability equivalent lottery method was used. The subject is provided a choice between uncertain and certain alternatives. The uncertain alternative provides the most preferred level of the alternative with some probability p and the least preferred with probability $(1-p)$. The certain alternative provides the evaluation measure level for which the corresponding utility will be determined. The p is varied until the subject is indifferent. The utility of the certain alternative is then equal to p . By considering each category between the least and most preferred levels, the utility for each category is determined. Again the subject rapidly provided indifference points for p .

2.4 ATR CS Alternatives. Data for most of the evaluation measures of the three alternatives, ATR 33, ATR 55, and ATR 89, were available from past MSTAR test data. As the MSTAR data did not include all evaluation measures, assumptions were made to permit analysis. These assumptions were:

- Developmental Costs are sunk. Value, $v(x)$, and utility, $u(x)$ are set equal to one in the model.
- Robustness for classify by class ($\% \Delta P_{CC}$) data was not collected. Data have $v(x) = u(x) = 1$.
- ATR CSs evaluated in the MSTAR program did not have the self-assessment capability: $v(x) = u(x) = 0$.
- No Classify by Class data was collected, so it was assumed that $P_{CC} = P_{ID}$. This provides a conservative estimate of P_{CC} .
- Assumed that costs for use time greater than 300 CPU seconds have $v(x) = u(x) = 0$ and were truncated to stay within elicited domain for this evaluation measure.
- The preference functions for the cost of redeployment of the system were elicited on a domain normalized for the Global Hawk system. The Global Hawk costs were assumed to be \$60k for this study.
- Operational costs were based on a \$10k price for a workstation with a three-year lifecycle.
- Distributions were assumed to be triangular.

- P_{DEC} for all ATR CSs is assumed to be 1 since MSTAR CSs must make a declaration decision.

Some data were fictionalized from their test values to mask competition sensitive information. Fictional data are representative and the results may be used to validate the methodological approach, but the results are not valid for ATR selection. The analysis would have to be repeated with true data for an ATR selection decision. The data for the alternatives is available through the COMPASE Center.

3.0 Results and Discussion. The expected value results for the alternatives examined with value functions (under conditions of uncertainty and certainty where the mean was treated as deterministic) and utility functions (under uncertainty) are provided in Table 3.1. The results for the value functions under certainty differ slightly from those employing the value function under uncertainty (note the CID profile for ATR 55). This illustrates that using the mean of a distribution and treating it deterministically and using a value function does not provide identical results as considering the stochastic nature of the problem within the analysis.

Table 3.1 ATR CS Expected Value and Expected Utility Results.

		ATR 33	ATR 55	ATR 89
Value Functions (Certainty)	CID	0.509	0.537	0.525
	ISR	0.497	0.531	0.497
Value Functions (Uncertainty)	CID	0.509	0.556	0.525
	ISR	0.497	0.531	0.497
Utility Functions (Uncertainty)	CID	0.572	0.507	0.518
	ISR	0.414	0.455	0.439

Comparing, by rank ordering the alternatives, the results of the value functions under uncertainty and the utility functions in Table 3.2, we see that the recommended alternative differs for the CID case. This indicates that the constructs of value and utility provide differing answers. As the problem involves uncertainty, the utility results are the appropriate choice. The two employment profiles provide differing recommendations. For CID missions, ATR 33 is the best choice. For ISR missions, ATR 55 is the preferred alternative. However ATR 89 was the second choice under both profiles. Should one ATR be required to perform both missions, ATR 89 may be the most appropriate choice. These results are depicted graphically in Figure 3.1. The risk profiles (as cumulative distribution functions [CDFs] are referred to in DA terminology) for the utility distributions for the CID and ISR profiles are provided in Figures 3.2 and 3.3. Figures 3.4 and 3.5 present tornado diagrams for the two profiles where the weights were varied by ten percent. The lack of color change (shading) in the bars indicates that the ATR choice is not sensitive to the weights elicited from the decision maker.

Table 3.2 Recommended ATR CS Alternative by Expected Value and Utility.

		ATR 33	ATR 55	ATR 89
Value Functions (Certainty)	CID	3	1	2
	ISR	2	1	2
Value Functions (Uncertainty)	CID	3	1	2
	ISR	2	1	2
Utility Functions (Uncertainty)	CID	1	3	2
	ISR	3	1	2

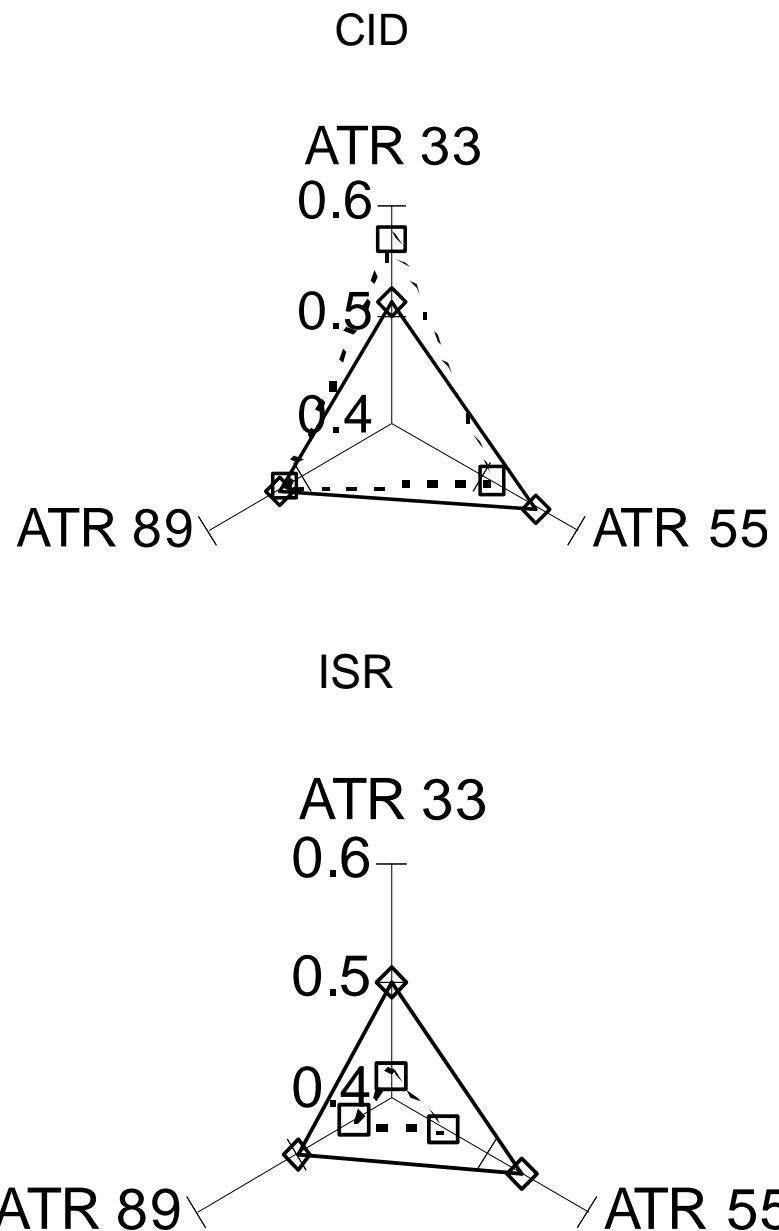


Figure 3.1 Radar Plots of ATR Value and Utility. Value results are represented by solid lines, utility by dashed. The Combat Identification (CID) results are above, the Intelligence/Surveillance/Reconnaissance (ISR) are below.

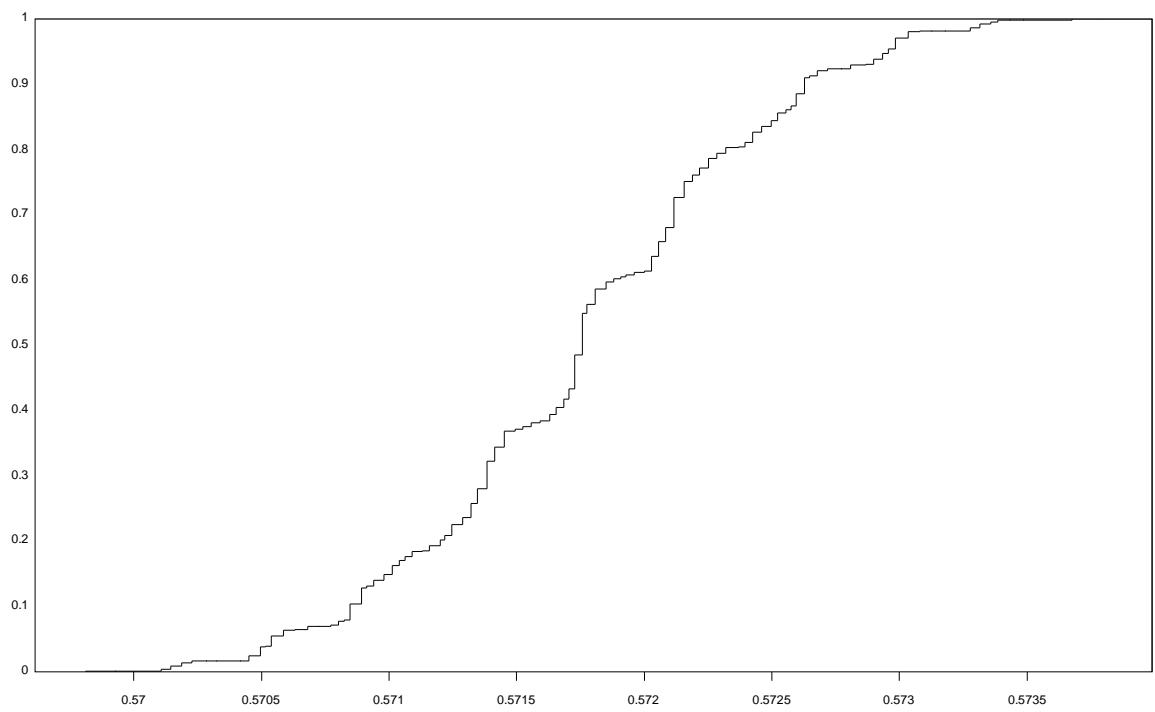


Figure 3.2 Risk Profile for CID.

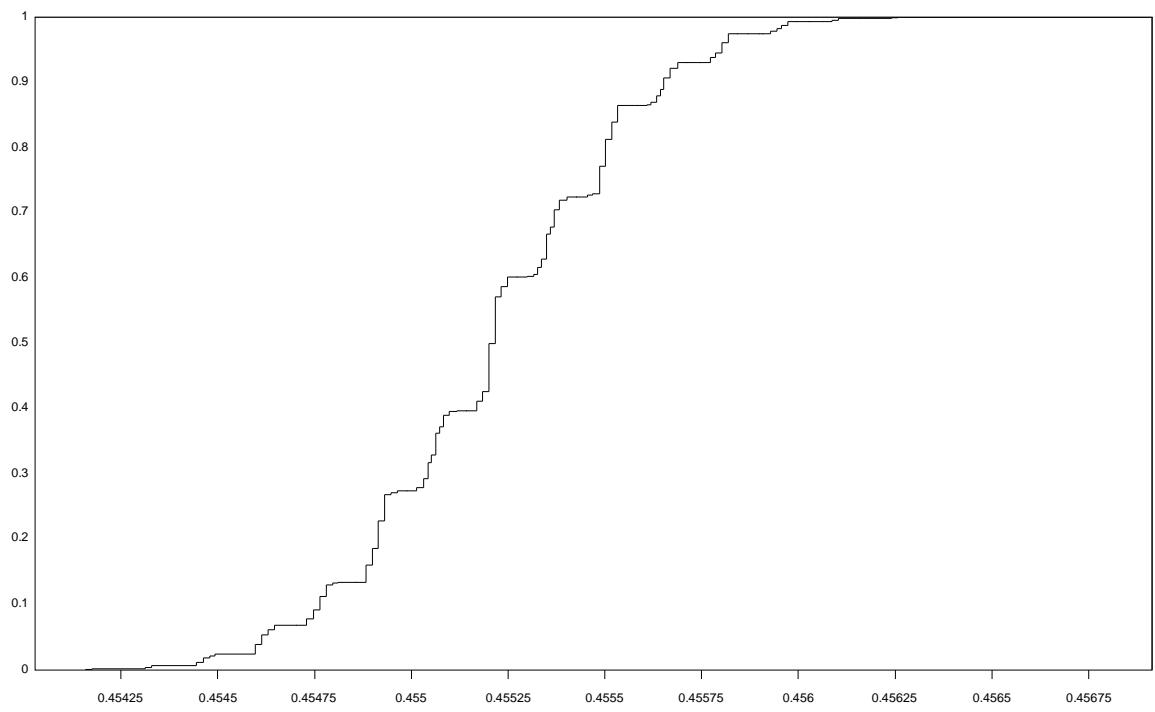


Figure 3.3 Risk Profile for ISR.

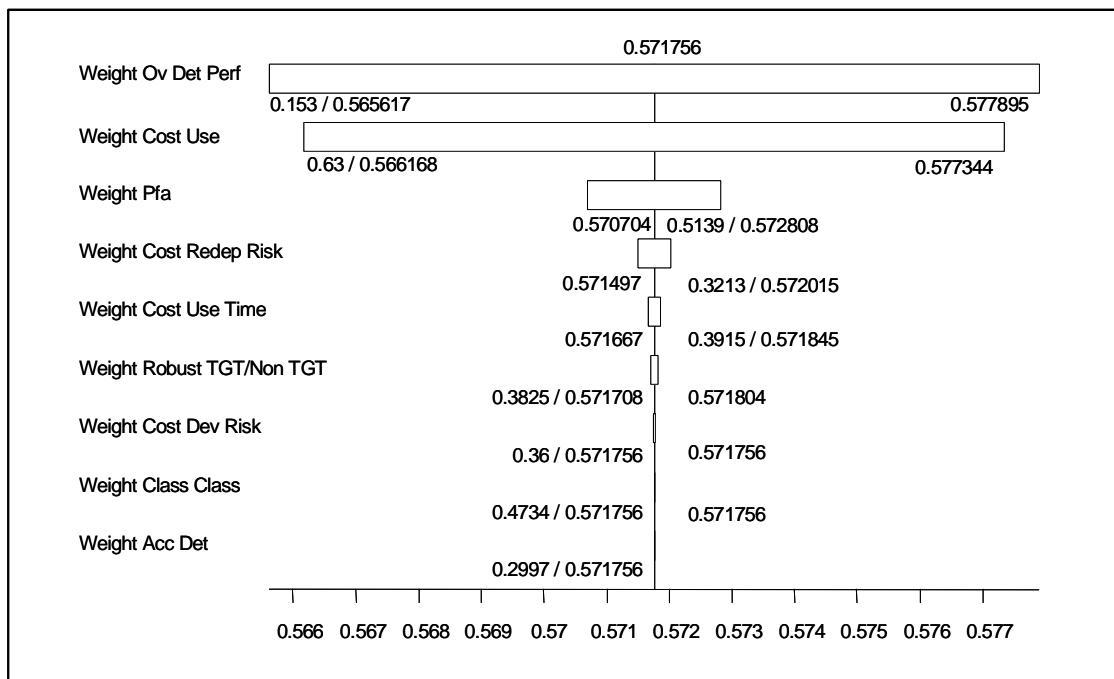


Figure 3.4 Sensitivity Analysis for CID Scenarios.

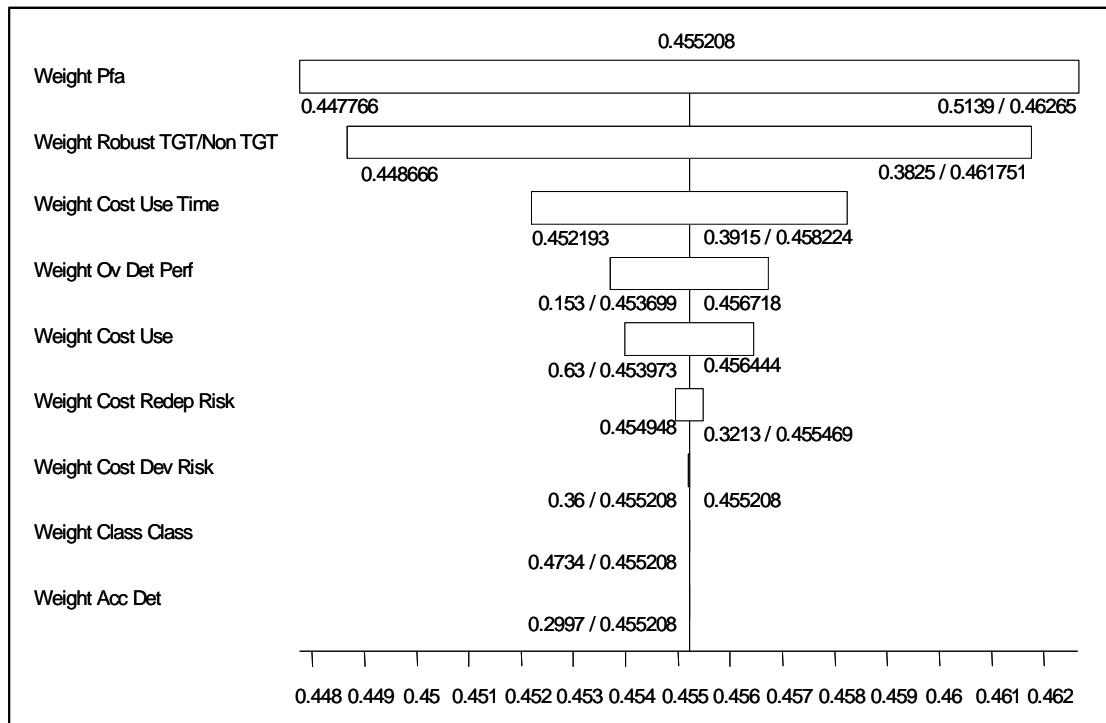


Figure 3.5 Sensitivity Analysis for ISR Scenarios.

The following figures illustrate that differing approaches, deterministic value, uncertain value and uncertain utility, provide differing answers. For the certain value approach, Figure 3.6 shows that for the CID employment profile, increasing cost of the ATR CS provides increasing value (benefit). For the ISR profile, ATR 33 dominates ATR 89. ATR 55 does provide improved value for an increased cost. These results hold for the uncertain value case, as illustrated in Figure 3.7. When utility is employed for the CID profile (Figure 3.8), ATR 33 dominates ATR 89, which in turn dominates ATR 55. For the ISR profile, increasing cost provides increasing utility (benefit). These interpretations agree with those of Table 3.2. Clearly both the mission profile and the DA methodology significantly affect the recommendation (answer).

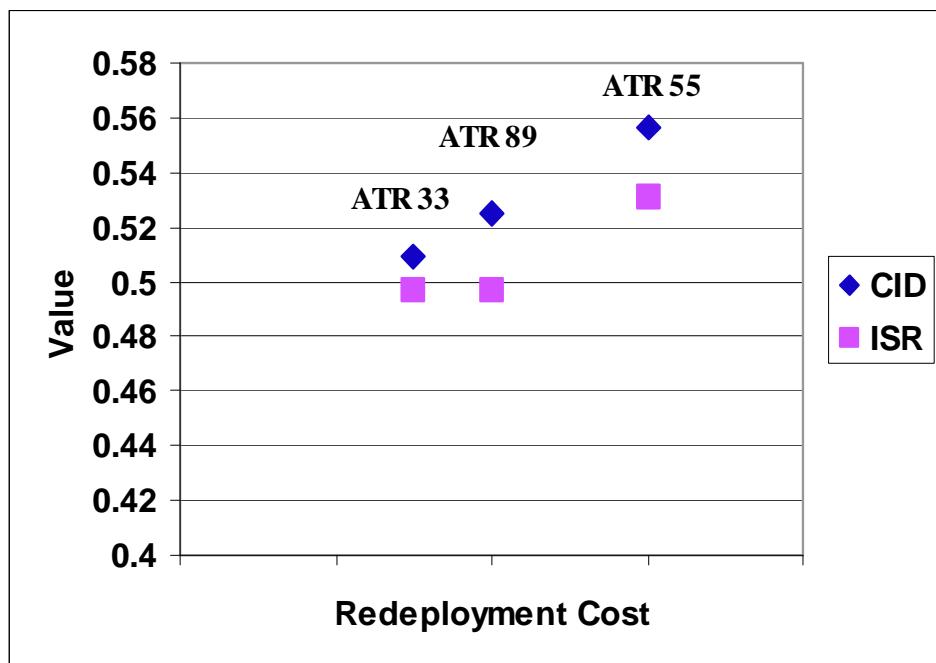


Figure 3.6 ATR Value (With Uncertainty) Versus Redeployment Cost.

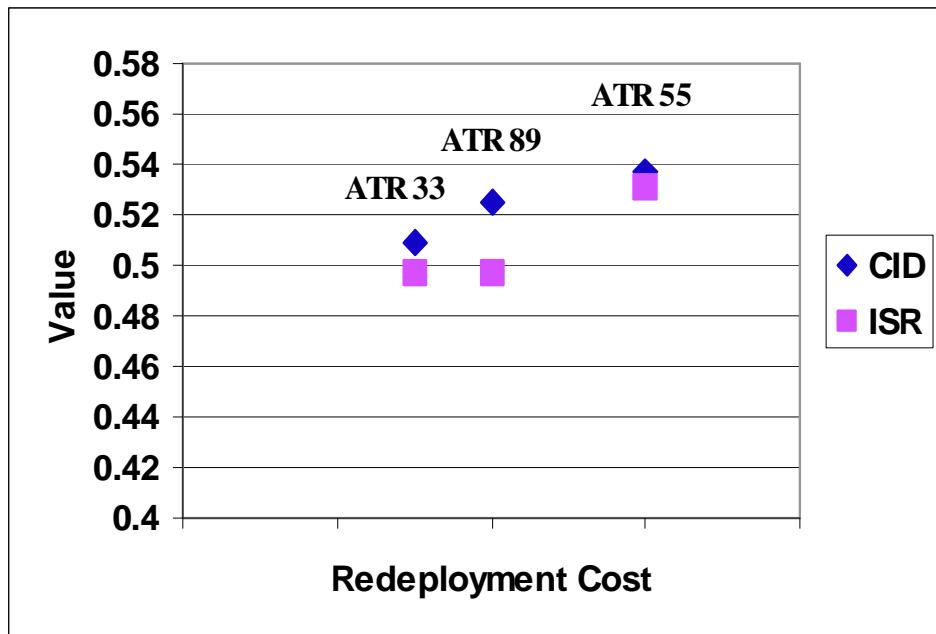


Figure 3.7 ATR Value (With Uncertainty) Versus Redeployment Cost.

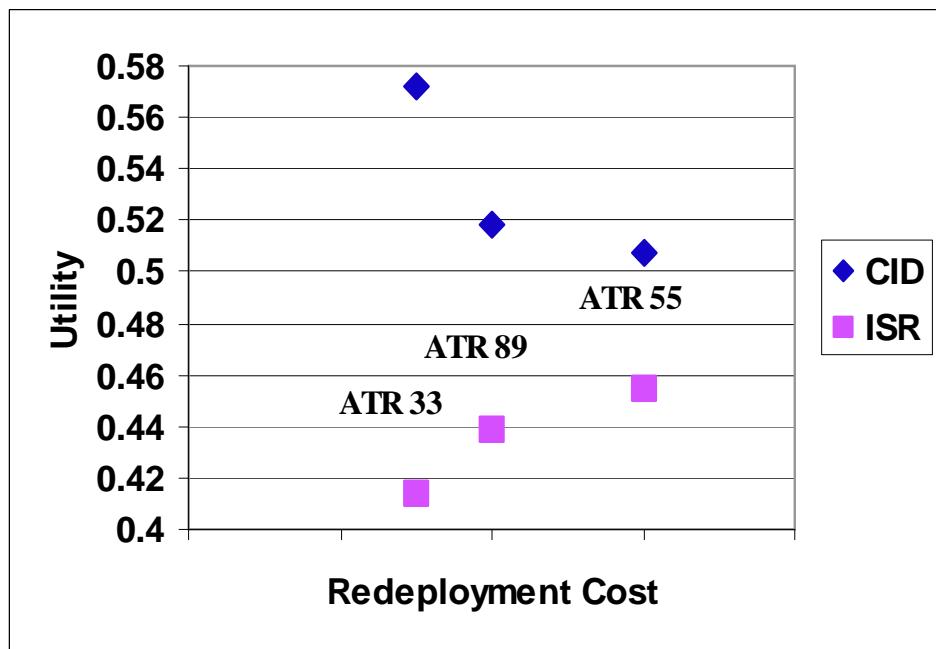


Figure 3.8 ATR Utility (With Uncertainty) Versus Redeployment Cost.

4.0 Conclusions. The conclusion can be drawn that the DA approach does indeed work.

The decision-maker's preferences are encapsulated in the framework, and the resultant

utility measure provides a defensible argument towards selecting a given ATR CS over another.

5.0 Recommendations. The authors recommend that these results should be applied towards a warfighter's perspective of the same problem. In a method parallel to the one detailed in this report, a DA framework should be constructed that provides a utility function measure for combat model results generated using similar ATR CS performance measures. The utility of the warfighter perspective may then be used in conjunction with the ATR technology developer's utility results in order to positively affect the manner in which ATR products are evaluated and judged.

It is recommended that the decision analysis approach should be incorporated into ATR research and development programs. Design of experiments and data collection efforts should serve to provide data to score alternatives in accordance with the value hierarchy. Assessment of ATR programs should include mission profile considerations. Decision analysis evaluations should be utility based unless all criteria are considered to be under conditions of certainty, which is unlikely.

References

- [1] Air Force Research Laboratory COMPASE Center, *ATR Performance Metrics*.
https://restricted.compase.vdl.afrl.af.mil/agri/acc_summer_2000/metrics.html
- [2] Alsing, S.G., *The Evaluation of Competing Classifiers*. Ph.D. dissertation, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, 2000.
- [3] Alsing, S. G., Kenneth W. Bauer, Jr., and John O. Miller, Evaluating Competing Classifiers Using a Multinomial Selection Procedure, *Proceedings of Artificial Neural Networks in Engineering (ANNIE) 2000 Conference*. St. Louis, MO, November 5-8, 2000.
- [4] Committee of Technology for Future Naval Forces. Technology for the United States Navy and Marine Corps, 2000-2035: Becoming a 21st-Century Force. Volume 2, Technology. http://www.nap.edu/html/tech_21st/t1.htm. Washington, DC: Naval Academy Press, 1997.
- [5] Ross, T. D., and Mossing, J. C., The MSTAR Evaluation Methodology. In *Proceedings of the International Society for Optical Engineering (SPIE)*, vol. 3721.

Appendix A – Elicitation Results

The results of the elicitation meetings between the authors and the COMPASE Center representative are summarized below. Elicitation sessions also included hand-drawn plots of the value functions, which are not presented here.

1. Robustness.
 - a. Top level weight: 0.2
 - b. Subvalues:
 - i. Detection Robustness
 1. Units: percent change in P_d .
 2. Domain: $[0,1]$.
 3. Weight: 0.425.
 4. Value Function: exponential fitting $\{(0,1), (25,0.2), (50,0)\}$.
 5. Utility Function: $\{(0,1), (3,0.75), (10,0.5), (10,0.25), (50,0)\}$.
 - ii. Identification Robustness
 1. Units: percent change in P_{id} .
 2. Domain: $[0,1]$.
 3. Weight: 0.275.
 4. Value Function: exponential fitting $\{(0,1), (25,0.2), (50,0)\}$.
 5. Utility Function: $\{(0,1), (3,0.75), (10,0.5), (10,0.25), (50,0)\}$.
 - iii. Classification Robustness
 1. Units: percent change in P_{cc} .
 2. Domain: $[0,1]$.
 3. Weight: 0.3.
 4. Value Function: exponential fitting $\{(0,1), (25,0.2), (50,0)\}$.
 5. Utility Function: $\{(0,1), (3,0.75), (10,0.5), (10,0.25), (50,0)\}$.
 2. Overall Detection Performance. Note that this value decomposes into either the “defined P_d ” subvalue or the “ROC” subvalue set, depending on which metric is used.
 - a. Top level weight: 0.17.
 - b. Subvalues:
 - i. Defined P_d .
 1. $FAR | P_d$. Note, $FAR | P_d = 0.9$ or $FAR | P_d = 0.5$, depending on mission profile.
 - a. Units: *occurrences / km²*.
 - b. Domain: $[0,1]$.

- c. Weight: 0.429.
- d. Value Function: exponential fitting $\{(0,1), (1,0.3), (1000,0)\}$.
- e. Utility Function: $\{(0,1), (3,0.75), (10,0.5), (50,0.25), (1000,0)\}$.

2. $P_{FA} | P_d$. Note, $P_{FA} | P_d = 0.9$ or $P_{FA} | P_d = 0.5$, depending on mission profile.

- a. Units: probability.
- b. Domain: $[0,1]$.
- c. Weight: 0.571.
- d. Value Function: linear fitting $\{(0,1), (1,0)\}$.
- e. Utility Function: $\{(0,1), (0.1,0.75), (0.3,0.5), (0.4,0.25), (1,0)\}$.

ii. ROC.

1. $ROC FAR$.
 - a. Units: area under ROC curve.
 - b. Domain: $[0,1000]$.
 - c. Weight: 0.429.
 - d. Value Function: exponential fitting $\{(0,0), (500,0.1), (1000,1)\}$.
 - e. Utility Function: $\{(0,0), (600,0.25), (750,0.5), (875,0.75), (1000,1)\}$.
2. $ROC P_{FA}$.
 - a. Units: normalized area under ROC curve.
 - b. Domain: $[0,1]$.
 - c. Weight: 0.571.
 - d. Value Function: exponential fitting $\{(0.5,0), (0.75,0.3), (1,1)\}$.
 - e. Utility Function: $\{(0.5,0), (0.8,0.25), (0.875,0.5), (0.938,0.75), (1,1)\}$.

3. Employment Concept.

- a. Units: categorical.
- b. Domain: {None, Poorly Defined, Moderately Defined, Strongly Defined, Well Defined}.
- c. Weight: 0.15.
- d. Value Function: {(None, 0), (Poorly Defined, 0.4), (Moderately Defined, 0.6), (Strongly Defined, 0.9), (Well Defined, 1)}.
- e. Utility Function: {(None, 0), (Poorly Defined, 0.1), (Moderately Defined, 0.5), (Strongly Defined, 0.9), (Well Defined, 1)}.

4. Declaration Ability.

- a. Units: probability.
- b. Domain: $[0.5,1]$.
- c. Weight: 0.13.

- d. Value Function: exponential fitting $\{(0,0), (0.5,0.7), (1,1)\}$.
- e. Utility Function: $\{(0,0), (0.25,0.25), (0.5,0.5), (0.75,0.75), (1,1)\}$.

5. Classification Ability.

- a. Top level weight: 0.11.
- b. Subvalues:
 - i. Correctly Classify by Type (P_{ID}).
 - 1. Units: probability.
 - 2. Domain: $[0.5,1]$.
 - 3. Weight: 0.474.
 - 4. Value Function: exponential fitting $\{(0.5,0), (0.75,0.3), (1,1)\}$.
 - 5. Utility Function: $\{(0.5,0), (0.6,0.25), (0.7,0.5), (0.9,0.75), (1,1)\}$.
 - ii. Correctly Classify by Class (P_{CC}).
 - 1. Units: probability.
 - 2. Domain: $[0,1]$.
 - 3. Weight: 0.526.
 - 4. Value Function: exponential fitting $\{(0.5,0), (0.75,0.2), (1,1)\}$.
 - 5. Utility Function: $\{(0.5,0), (0.6,0.25), (0.7,0.5), (0.9,0.75), (1,1)\}$.

6. Cost

- a. Top level weight: 0.1.
- b. Subvalues:
 - i. Development
 - ii. Weight: 0.01.
 - 1. Money.
 - a. Units: \$M.
 - b. Domain: $[0,10]$.
 - c. Weight: 0.24.
 - d. Value Function: exponential fitting $\{(0,1), (5,0.3), (10,0)\}$.
 - e. Utility Function: $\{(0,1), (2.5,0.75), (5,0.5), (7.5,0.25), (10,0)\}$.
 - 2. Time.
 - a. Units: months.
 - b. Domain: $[0,18]$.
 - c. Weight: 0.2.
 - d. Value Function: exponential fitting $\{(0,1), (9,0.3), (18,0)\}$.
 - e. Utility Function: $\{(0,1), (4.5,0.75), (9,0.5), (13.5,0.25), (18,0)\}$.
 - 3. Expertise
 - a. Units: categorical.

- b. Domain: {Technical Training, BS in Engineering, Graduate-level Engineer, Multi-location subject matter experts (SME), Single Site SME}.
- c. Weight: 0.16.
- d. Value Function: {(Technical Training, 1), (BS in Engineering, 0.8), (Graduate-level Engineer, 0.4), (Multi-location subject matter experts (SME), 0.2), (Single Site SME, 0)}.
- e. Utility Function: {(Technical Training, 1), (BS in Engineering, 0.9), (Graduate-level Engineer, 0.8), (Multi-location subject matter experts (SME), 0.7), (Single Site SME, 0)}.

4. Risk

- a. Units: categorical.
- b. Domain: Low, Medium, and High.
- c. Weight: 0.4.
- d. Value Function: exponential fitting: {(Low, 1), (Medium, 0.5), (High, 0)}.
- e. Utility Function: {(Low, 1), (Medium, 0.8), (High, 0)}.

iii. Redeployment.

iv. Weight: 0.29.

1. Money.

- a. Units: normalized on Global Hawk.
- b. Domain: $[0,1]$.
- c. Weight: 0.143.
- d. Value Function: linear, $\{(0,1), (1/3,0), (1,0)\}$.
- e. Utility Function: $\{(0,1), (0.15,0.75), (0.333,0.5), (0.9999,0.25), (1,0)\}$.

2. Time.

- a. Units: days.
- b. Domain: $[0,90]$.
- c. Weight: 0.179.
- d. Value Function: exponential fitting $\{(0,1), (45,0.2), (90,0)\}$.
- e. Utility Function: $\{(0,1), (15,0.75), (30,0.5), (50,0.25), (90,0)\}$.

3. Expertise

- a. Units: categorical.
- b. Domain: {Technical Training, BS in Engineering, Graduate-level Engineer, Multi-location subject matter experts (SME), Single Site SME}.
- c. Weight: 0.321.
- d. Value Function: {(Technical Training, 1), (BS in Engineering, 0.95), (Graduate-level Engineer, 0.9),

(Multi-location subject matter experts (SME), 0.5),
 (Single Site SME, 0)}.

e. Utility Function: {(Technical Training, 1), (BS in Engineering, 0.95), (Graduate-level Engineer, 0.9),
 (Multi-location subject matter experts (SME), 0.8),
 (Single Site SME, 0)}.

4. Risk

- a. Units: categorical.
- b. Domain: Low, Medium, and High.
- c. Weight: 0.357.
- d. Value Function: exponential fitting: {(Low, 1),
 (Medium, 0.5), (High, 0)}.
- e. Utility Function: {(Low, 1), (Medium, 0.2), (High, 0)}.

v. Use.

vi. Weight: 0.7.

1. Money.

- a. Units: normalized on Global Hawk.
- b. Domain: [0,2].
- c. Weight: 0.217.
- d. Value Function: exponential fitting
 $\{(0,1), (1,0.2), (2,0)\}$.
- e. Utility Function:
 $\{(0,1), (0.5,0.75), (1,0.5), (1.5,0.25), (2,0)\}$.

2. Time.

- a. Units: minutes.
- b. Domain: [0,5].
- c. Weight: 0.435.
- d. Value Function: exponential fitting
 $\{(0,1), (2 : 30,0.3), (5,0)\}$.
- e. Utility Function:
 $\{(0,1), (1,0.75), (2,0.5), (3,0.25), (5,0)\}$.

3. Expertise

- a. Units: categorical.
- b. Domain: {Technical Training, BS in Engineering, Graduate-level Engineer, Multi-location subject matter experts (SME), Single Site SME}.
- c. Weight: 0.348.
- d. Value Function: {(Technical Training, 1), (BS in Engineering, 0.4), (Graduate-level Engineer, 0.3),
 (Multi-location subject matter experts (SME), 0.1),
 (Single Site SME, 0)}.
- e. Utility Function: {(Technical Training, 1), (BS in Engineering, 0.2), (Graduate-level Engineer, 0.1),

(Multi-location subject matter experts (SME), 0.05),
(Single Site SME, 0)}.

7. Self-Assessed Accuracy.

a. Es-Pd

- i. Units:
- ii. Domain: [0,1].
- iii. Weight: 0.333.
- iv. Value Function: exponential fitting $\{(0,1), (0.5,0.7), (1,0)\}$.
- v. Utility Function: $\{(0,1), (0.25,0.75), (0.5,0.5), (0.75,0.25), (1,0)\}$.

b. Es-Pid

- i. Units:
- ii. Domain: [0,1].
- iii. Weight: 0.333.
- iv. Value Function: exponential fitting $\{(0,1), (0.5,0.7), (1,0)\}$.
- v. Utility Function: $\{(0,1), (0.25,0.75), (0.5,0.5), (0.75,0.25), (1,0)\}$.

c. Es-Pcc

- i. Units:
- ii. Domain: [0,1].
- iii. Weight: 0.333.
- iv. Value Function: exponential fitting $\{(0,1), (0.5,0.7), (1,0)\}$.
- v. Utility Function: $\{(0,1), (0.25,0.75), (0.5,0.5), (0.75,0.25), (1,0)\}$.

List of Acronyms

AFRL	Air Force Research Laboratory
AGRI	Air-to-Ground Radar Imaging
ATR	Automatic Target Recognition
CID	Combat Identification
COMPASE Center	Comprehensive ATR Scientific Evaluation Center
CS	Classification System
DA	Decision Analysis
ISR	Intelligence/Surveillance/Reconnaissance
MBT	Main Battle Tank
MSTAR	Moving and Stationary Target Acquisition and Recognition
ROC	Receiver Operating Characteristic
SAR	synthetic aperture radar
TGT	target

REPORT DOCUMENTATION PAGE				<i>Form Approved OMB No. 074-0188</i>
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>				
1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE		3. DATES COVERED (From – To)	
29-04-2002	Technical Report		JUL 2001-APR 2002	
4. TITLE AND SUBTITLE Application of Decision Analysis to Automatic Target Recognition Programmatic Decisions			5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER 5d. PROJECT NUMBER ENR# 2001-018	
6. AUTHOR(S) William K. Klimack, Col, USA Christopher B. Bassham, Capt, USAF Dr. Kenneth W. Bauer, Jr., Professor			5e. TASK NUMBER 5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) AFIT/ENS 2950 P St, Building 640 Dept. of Operational Sciences Air Force Institute of Technology WPAFB OH 45433-7765			8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/EN-TR-02-06	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/ Sensors Directorate ATTN: James D. Leonard 2241 Avionics Cl Building 620 Rm N3-X19 WPAFB OH 45433			10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT The purpose of this research is to demonstrate the application of decision analysis (DA) techniques to the decisions made throughout the lifecycle of Automatic Target Recognition (ATR) technology development. This work is accomplished in the hopes of improving the means by which ATR technologies are evaluated. The first step in this research was to create a flexible decision analysis framework that could be applied to a variety of decisions across several different ATR programs evaluated by the Comprehensive ATR Scientific Evaluation (COMPASE) Center. For the purposes of this research, a single COMPASE Center representative provided the value, utility, and preference functions for the DA framework. The DA framework employs performance measures collected during ATR classification system (CS) testing to calculate value and utility scores. The authors gathered data from the Moving and Stationary Target Acquisition and Recognition (MSTAR) program to demonstrate how the decision framework could be used to evaluate three different ATR CSs. A decision-maker may use the resultant scores to gain insight into any of the decisions that occur throughout the lifecycle of ATR technologies. Additionally, a means of evaluating ATR CS self-assessment ability is presented.				
15. SUBJECT TERMS Automatic Target Recognition, Decision Analysis, Competing Classifiers, Measures of Performance, Utility, Value, Comparison, Value Hierarchy, self-assessment, evaluation, ROC curve, MSTAR				
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 38	19a. NAME OF RESPONSIBLE PERSON Dr. Kenneth W. Bauer, Jr., Professor
a. REPORT U b. ABSTRACT U c. THIS PAGE U				19b. TELEPHONE NUMBER (Include area code) (937) 255-6565, ext 4328